



Estimation of Population Parameters in Genetic Polymorphism Equilibrium by Overdominance Model*

Zhou Hougao

(College of Agronomy, Guangxi University, Nanning, 530005)

Zhu Jun Yan Xinfu

(Department of Agronomy, Zhejiang University, Hangzhou 310029)

Abstract: By mathematical approaches, the maintenance mechanism of genetic polymorphism in population as well as the conditions of equilibrium was demonstrated based on over-dominance model. It was found that principles derived from two-allelic case had limitation. The principles used in many textbooks for estimating allelic frequencies in equilibrium were special ones that only suit two-allelic case. The extension to multi-allelic case was performed in alternative approach and a set of formulas were developed in this paper. The relationship of allele number (n) at one locus to average fitness (\bar{w}), genetic load (L), heterozygote frequency (H_e) and homozygote frequency (H_{om}) in population were discussed.

Key words: Overdominance model; Genetic polymorphism equilibrium; Allelic frequency; Average fitness; Genetic load

Document code: A **Articld ID:** 1001-9626(2000)01-0021-08

1 Introduction

Existence of genetic polymorphism in natural plant and animal populations has been generally recognized. Two theories, neutralist and balance theory, were proposed to explain the maintenance mechanism of polymorphism. The neutralist theory or neo-classical theory believes that the polymorphism is due to the accumulation of unselected mutations and modulated by random sampling events (Kimura 1979; Merrel 1981). Some mutational theories were proposed, such as the "infinite alleles" model of Kimura (1968) and the "ladder model" of Ohta and Kimura (1973) as well as and the "stepwise model" of Kimura and Ohta (1978). The balance theory (selection theory) explains the polymorphism to be the result of the balance of selective forces. When genetic polymorphism is in equilibrium, population genetic parameters such as allelic frequency and genetic load which are obtained by the overdominance model for two alleles at a locus are a function of selection coefficients of homozygotes (Wright 1953). Thus, a very important conclusion is that the overdominance of heterozygotes can make a stable balance of genetic polymorphism in

Received date: 1998-03-19

* **Foundation item:** The work was supported by the National Natural Science Foundation of China

population and the balance depends only upon the selection coefficient of the heterozygotes. Some people assumed that the finding obtained in two-allelic case seemed to be suitable for genetic polymorphic equilibrium in the case of more than two alleles at one locus (Ford 1975; Kuo 1993). Then, many researchers (Li 1953; Kimura 1956; Mandel 1959; Lewontin 1978) tried to extend the conclusion to the case of multi-alleles at one locus and to estimate the population genetic parameters in genetic polymorphic equilibrium using mathematical formulas. However, since their formulas in matrix form were quite complex, the direct relationship between the genetic parameters and genetic polymorphic equilibrium of population could not be consequently determined in explicit terms. Lewontin (1978) put forward some conjecture on the related condition for maintaining gene equilibrium and genetic parameters in multi-allelic equilibrium in elicitation method. From his complex model, he drew an inference of genetic polymorphic equilibrium that n alleles were equal in frequencies i.e $p = 1/n$, and there was little variation in fitness among heterozygotes. But the inference has not been proved in mathematics and theory. In this paper, based on the overdominance model, we proposed some extension to multiallelic case from two-allelic case, developed some intuitive and easily interpretative formulae for estimating genetic parameters of population in genetic polymorphic equilibrium, and determined the relationships between population genetic parameters and allelic number for the first time, and proved Lewontin's conjecture (1978).

2 Genetic Polymorphism Equilibrium in Two-allelic Case

The topic has been concerned by many geneticists (Mandel 1959; Li 1953; Lewontin 1958, 1978; Kimura 1956; Wallace 1981; Spiess 1983; Hart 1980; Merrell 1981). Overdominance model believes that the fitness of heterozygotes is superior to that of homozygote of the two alleles from heterozygote. Selection will result in the decrease of homozygote frequency and the increase of heterozygote frequency. The decrease of homozygotes A_1A_1 and A_2A_2 causes the decrease of frequencies of alleles A_1 and A_2 respectively, but the increase of heterozygotes A_1A_2 causes the increase of frequencies of alleles A_1 and A_2 . The frequencies of the two alleles A_1 and A_2 will reach stable equilibrium at balance point due to the opposite actions of selection.

Set p = frequency of allele A_1 , q = frequency of allele A_2 , s_1 = selective index of homozygote A_1A_1 , s_2 = selective index of homozygote A_2A_2 , and $s_1 > 0$, $s_2 > 0$, $p \neq 0$, $q \neq 0$. Then the changes of genotypic frequencies in population after selection can be calculated and listed in table 1.

Table 1 Analysis of Selectively Superior Heterozygotes in Two-allelic Case

Genotype	A_1A_1	A_1A_2	A_2A_2	total
Initial frequency	p^2	$2pq$	q^2	1
Selective index	s_1	0	s_2	
Fitness	$1 - s_1$	1	$1 - s_2$	
Final frequency	$p^2(1 - s_1)$	$2pq$	$q^2(1 - s_2)$	$1 - s_1p^2 - s_2q^2$

Average fitness (\bar{w}) of population after selection is

$$\bar{w} = p^2(1 - s_1) + 2pq + q^2(1 - s_2) = 1 - s_1p^2 - s_2q^2. \quad (1)$$

The frequency of allele after one generation is

$$q_1 = \frac{q(1 - s_2q)}{1 - s_1p^2 - s_2q^2}. \quad (2)$$

The change of allelic frequency after one generation is

$$\Delta q = q_1 - q = \frac{pq(s_1p - s_2q)}{1 - s_1p^2 - s_2q^2} \quad (3)$$

i.e.

$$\Delta q = \frac{q(s_1p^2 + s_2q^2 - s_2q)}{1 - s_1p^2 - s_2q^2}. \quad (4)$$

According to (3), if the equilibrium is established, allelic frequency changes no longer and the necessary condition for $\Delta q = 0$ is $s_1p - s_2q = 0$. Derivation $s_2q = s_1p = s_1(1 - q) = s_1 - s_1q$, we can get

$$\hat{q} = \frac{s_1}{s_1 + s_2}, \quad \hat{p} = \frac{s_2}{s_1 + s_2}. \quad (5)$$

Equations in (5) are often used in textbooks, they indicate that 1) the equilibrium established under the selective superiority of heterozygote individuals is a stable equilibrium, it can approach to equilibrium from either of side and tends to be restored if the population is perturbed; 2) the equilibrium depends on the selective indices of homozygotes and has no relation with initial allele frequencies. The principles derived from the two-allelic case are special ones. Since equations in (5) are feasible for two-allelic case only, extension of them to three or more alleles may be misleading. Under what conditions does the three or more allelic system at a locus achieve stable equilibrium? What are the frequencies of alleles at stationary state? These problems had been partly answered by Kimura (1959) and Lewontin (1958, 1978). In next section we shall give some intuitive and easily interpretative formulas under the hypothesis of overdominance.

Derivation: from (4), the necessary condition for setting $\Delta q = 0$ is $s_1p^2 + s_2q^2 - s_2q = 0$, thus

$$\hat{q} = (s_1p^2 + s_2q^2)/s_2, \quad (6)$$

$$\hat{q} = (s_1p^2 + s_2q^2)/s_1. \quad (7)$$

The equations (6) and (7) reveal that the allele frequency in equilibrium is correlated negatively with selective index of its homozygous individuals and positively correlated with selection quantity ($s_1p^2 + s_2q^2$) of all homozygotes. For a given set of starting allele frequencies, the selective indices of homozygotes play an important part in determining equilibrium allele frequencies.

3 Genetic Polymorphism Equilibrium in Three-allelic Case

Set frequencies of alleles A_1 , A_2 and A_3 to be p , q and r , respectively, where p , q and r do not equal zero with selective indices of their homozygotes being s_1 , s_2 and s_3 , respectively, and selective indices of all heterozygotes equalling zero.

Table 2 Analysis of Selectively Superior Heterozygotes in Three-allelic Case

Genotyp.	A_1A_1	A_2A_2	A_3A_3	A_1A_2	A_1A_3	A_2A_3	Total
Initial freq.	p^2	q^2	r^2	$2pq$	$2pr$	$2qr$	1
Select.ind.	s_1	s_2	s_3	0	0	0	
Fitness	$1 - s_1$	$1 - s_2$	$1 - s_3$	1	1	1	
Final Freq.	$p^2(1 - s_1)$	$q^2(1 - s_2)$	$r^2(1 - s_3)$	$2pq$	$2pr$	$2qr$	$1 - s_1p^2 - s_2q^2 - s_3r^2$

Average fitness of populations is

$$\bar{w} = p^2(1 - s_1) + q^2(1 - s_2) + r^2(1 - s_3) + 2pq + 2pr + 2qr = 1 - s_1p^2 - s_2q^2 - s_3r^2. \quad (8)$$

The frequency of allele after one generation is

$$q_1 = \frac{q^2(1 - s_2) + pq + qr}{1 - s_1p^2 - s_2q^2 - s_3r^2} = \frac{q[(1 - s_2)q + p + r]}{1 - s_1p^2 - s_2q^2 - s_3r^2} = \frac{(1 - s_2)q}{1 - s_1p^2 - s_2q^2 - s_3r^2}, \quad (9)$$

the change of alleles frequency is

$$\Delta q = q_1 - q = \frac{q[s_1p^2 + s_3r^2 - s_2q(1 - q)]}{1 - s_1p^2 - s_2q^2 - s_3r^2} = \frac{q[p(s_1p - s_2q) + r(s_3r - s_2q)]}{1 - s_1p^2 - s_2q^2 - s_3r^2}. \quad (10)$$

We also get

$$\Delta p = \frac{p[q(s_2q - s_3r) + r(s_3r - s_1p)]}{1 - s_1p^2 - s_2q^2 - s_3r^2}, \quad (11)$$

$$\Delta r = \frac{r[p(s_1p - s_3r) + q(s_2q - s_3r)]}{1 - s_1p^2 - s_2q^2 - s_3r^2}. \quad (12)$$

From equation (10), the necessary condition for $\Delta q = 0$ is

$$p(s_1p - s_2q) + r(s_3r - s_2q) = 0. \quad (13)$$

Derivation : from (13), we can obtain $s_1p^2 + s_3r^2 = s_2pq + s_2rq = s_2q(p + r) = s_2q(1 - q) = s_2q - s_2q^2$, $s_2q = s_1p^2 + s_2q^2 + s_3r^2$. Thus

$$\hat{q} = (s_1p^2 + s_2q^2 + s_3r^2)/s_2. \quad (14)$$

From equations (11) and (12), we can also get

$$\hat{p} = (s_1p^2 + s_2q^2 + s_3r^2)/s_1, \quad (15)$$

$$\hat{r} = (s_1p^2 + s_2q^2 + s_3r^2)/s_3. \quad (16)$$

Equations (14)–(16) are similar to equations (6) and (7), the only difference is that the latter have more one term, s_3r^2 , in numerator. The relationships of allele frequencies in equilibrium to selective indices of their homozygotes are not different from two-allelic case.

4 Genetic Polymorphism Equilibrium in N -Allelic Case

We suppose there are n alleles at one locus, $A_1, A_2, A_3, \dots, A_i, \dots, A_n$, set $s_1, s_2, s_3, \dots, s_i$,

\dots, s_n to be selective indices of their homozygotes and $f_1, f_2, f_3, \dots, f_i, \dots, f_n$ to be the initial frequencies of alleles.

According to the same philosophy in two-and three-allelic case, we can get average fitness of population:

$$\bar{w} = 1 - s_1 f_1^2 - s_2 f_2^2 - \dots - s_i f_i^2 - \dots - s_n f_n^2. \tag{17}$$

The frequencies of alleles after one generation is

$$f'_j = \frac{(1 - s_j f_j) f_j}{1 - \sum_{i=1}^n (s_i f_i^2)}, \tag{18}$$

the change of allele frequency is

$$\Delta f_j = f'_j - f_j = \frac{f_j \sum_{i=1}^n f_i (s_i f_i - s_j f_j)}{1 - \sum_{i=1}^n s_i f_i^2}, \quad i \neq j, \quad j = 1, 2, 3, \dots, n. \tag{19}$$

We can get n equations of Δf_j here.

From equation (19), the necessary condition for $\Delta f_j = 0$ is

$$\sum_{i=1}^n f_i (s_i f_i - s_j f_j) = 0, \quad j = 1, 2, 3, \dots, n, \quad i \neq j, \tag{20}$$

we can get n equations also from equation (20) to estimate the frequency of allele j in equilibrium:

$$\hat{f}_j = (s_1 f_1^2 + s_2 f_2^2 + \dots + s_n f_n^2) / s_j = \sum_{i=1}^n s_i f_i^2 / s_j, \quad j = 1, 2, \dots, n. \tag{21}$$

Equation (21) is a general formula and an extension of equations (6), (7), (14), (15) and (16) in two-and three-allelic cases to n -allelic case.

5 The Number of Alleles and Frequencies of Homozygotes and Heterozygotes

In three-allelic case, if the population is in equilibrium, $\hat{p} + \hat{q} + \hat{r} = 1$, from equations (14)–(16), we can obtain

$$(1/s_1 + 1/s_2 + 1/s_3)(s_1 p^2 + s_2 q^2 + s_3 r^2) = 1, \tag{22}$$

If $s_1 = s_2 = s_3 = s$, from equation (22), we can also obtain

$$p^2 + q^2 + r^2 = 1/3. \tag{23}$$

Equation (23) indicates that if selective indices of all homozygotes are the same, then the frequency is 1/3 for all homozygotes and 2/3 for all heterozygotes. These results can be easily

extended to the n -allelic case. The frequency of all homozygotes (H_{om}) is

$$H_{om} = \sum_{j=1}^n f_j^2 = f_1^2 + f_2^2 + f_3^2 + \cdots + f_n^2 = \frac{1}{n}, \quad (24)$$

and the frequency of all heterozygotes (H_e) is

$$H_e = \sum_{i=1}^n \sum_{j=1}^n f_i f_j = \frac{n-1}{n} = 1 - \frac{1}{n}. \quad (25)$$

From equations (24) and (25), based on overdominance model, we can conclude that 1) if selective indices of all homozygotes are the same and the population is in equilibrium, the frequencies of homozygotes will decrease and the frequencies of heterozygotes will increase with the increase of allele number at a locus; 2) frequency decrease of one kind of homozygotes will cause the frequency decrease of the allele from the homozygote, while the increase of heterozygote frequency will cause increase of the frequency of alleles in heterozygotes. The opposite forces will drive the frequency of the allele to an equilibrium; 3) high frequencies of heterozygotes provide the opportunity for many alleles to coexist in natural population and this is the reason why abundant genetic variation and polymorphism exists in population.

6 The Number of Alleles and Average Fitness of Population

When a population genetic polymorphism achieves equilibrium, the average fitness of the population is maximised (Merrel 1981), according to equation (17), the average fitness of population is

$$\bar{w} = 1 - s_1 f_1^2 - s_2 f_2^2 - \cdots - s_i f_i^2 - \cdots - s_n f_n^2 = 1 - (s_1 f_1^2 + s_2 f_2^2 + \cdots + s_i f_i^2 + \cdots + s_n f_n^2), \quad (26)$$

If $s_1 = s_2 = s_3 = \cdots = s_n$, from equation (24) and (26) we can obtain

$$\bar{w} = 1 - s(f_1^2 + f_2^2 + \cdots + f_i^2 + \cdots + f_n^2) = 1 - s/n. \quad (27)$$

Equation (27) indicates that 1) the average fitness of population in equilibrium will increase with the increase of allele number at one locus; 2) average fitness is negatively correlated with selective indices of homozygotes; 3) maintenance of genetic variation in population is benefit to promotion of its fitness and this may be another reason why there is plentiful genetic polymorphism in population.

7 The Number of Alleles and Genetic Load of Population

In two-allelic case, the genetic load of population L is

$$L = s_1 p^2 + s_2 q^2. \quad (28)$$

Derivation : from equation (5) (Kuo 1993), we can get

$$L = s_1 \left(\frac{s_2}{s_1 + s_2} \right)^2 + s_2 \left(\frac{s_1}{s_1 + s_2} \right)^2 = \frac{s_1 s_2}{s_1 + s_2}. \quad (29)$$

In equilibrium, $s_1p = s_2q$, thus $s_1 = s_2q/p$; $s_2 = s_1p/q$. From (29), we get

$$L = s_2q, \quad (30)$$

$$L = s_1p. \quad (31)$$

The equations (30) and (31) are valid only in two-allelic case. In order to extend the two-allelic case to the n -allelic case, we can derived from equations (6) and (7) in the two-allelic case,

$$L = s_1p^2 + s_2q^2 = \left(\frac{1}{s_1} + \frac{1}{s_2}\right)(s_1p^2 + s_2q^2)^2. \quad (32)$$

From (17) and (21), we can extend the equation (32) in two-allelic case to n -allelic case:

$$\begin{aligned} L &= s_1f_1^2 + s_2f_2^2 + \cdots + s_if_i^2 + \cdots + s_nf_n^2 \\ &= \left(\frac{1}{s_1} + \frac{1}{s_2} + \cdots + \frac{1}{s_n}\right)(s_1f_1^2 + s_2f_2^2 + \cdots + s_nf_n^2)^2 \\ &= \left(\sum_{j=1}^n \frac{1}{s_j}\right)\left(\sum_{j=1}^n s_jf_j^2\right)^2. \end{aligned} \quad (33)$$

If $s_1 = s_2 = \cdots = s_n = s$,

$$L = \frac{s^2n}{s}(f_1^2 + f_2^2 + \cdots + f_i^2 + \cdots + f_n^2)^2 = ns\left(\sum_{j=1}^n f_j^2\right)^2. \quad (34)$$

According to (24), we have

$$L = s/n. \quad (35)$$

Equation (35) reveals that the genetic load of population in equilibrium will decrease with the increase of number of alleles at a locus and positively correlates with selective indices of homozygotes. The equation (35) is the same in genetic explanation as equation (27).

8 Numerical Relationships among Frequencies of Alleles

In three-allelic case, if $s_1 = s_2 = s_3$, then we have

$$\hat{p} = (s_1p^2 + s_2q^2 + s_3r^2)/s_1 = p^2 + q^2 + r^2, \quad (36)$$

$$\hat{q} = (s_1p^2 + s_2q^2 + s_3r^2)/s_2 = p^2 + q^2 + r^2, \quad (37)$$

and

$$\hat{r} = (sp^2 + sq^2 + sr^2)/s = p^2 + q^2 + r^2. \quad (38)$$

Thus, in equilibrium, $\hat{p} = \hat{q} = \hat{r} = 1/3$.

In extension to n -allelic case from equation (24), we can get

$$\hat{f}_1 = \hat{f}_2 = \hat{f}_3 = \cdots = \hat{f}_n = 1/n. \quad (39)$$

Equation (39) indicates that 1) if selective indices of all homozygote are the same, the frequencies of all alleles are identical, equalling $1/n$; 2) when frequencies of all alleles are equal,

the average fitness of population is maximised. These results confirm the Lewontin's conjectures (1978) theoretically and mathematically. Using both numerical and analytical approach, Lewontin put forward the following conjectures that multiple alleles could be maintained if, in addition to heterosis, there was very little variation in fitness among heterozygotes and that the allelic frequency distribution and equilibrium would be very uniform, with all alleles very close to equal frequency $p = 1/n$.

[References]

- [1] Ford E B. *Ecological Genetics* (4th ed)[M]. London: Chapman and Hall, 1975. 23-66.
- [2] Hartel DL. *Principles of Population Genetics*[M]. Sunderland: MA Sinauer, 1980. 75-84.
- [3] Kimura M. Rules for testing stability of a selective polymorphism[J]. *Proc Natl Acad US*, 1956, **42**(2):336-340.
- [4] Kimura M. The neutral theory of molecular evolution[J]. *Sci Amer*, 1979, **241**(5):98-126.
- [5] Kimura M, Ohta T. Stepwise mutation model and distribution of allelic frequencies in a finite population[J]. *Proc Natl Acad US*, 1978, **75**(16):2868-2872.
- [6] Kojima K, Yarbrough K M. Frequency dependent selection at the Esterase-6 locus in *Drosophila melanogaster*[J]. *Proc Nat Acad Sci*, 1967, **57**(3):645-649.
- [7] Kuo P Z. *The Introduction of Population Genetics*[M]. Beijing: Agricultural Press of China, 1993. 45-80.
- [8] Lewontin R C. A general method for investigating the equilibrium of gene frequencies in a population[J]. *Genetics*, 1958, **43**(3):419-434.
- [9] Lewontin R C, Ginzburg L R, Tuljapurkar S D. Heterosis as an explanation for large amount of genic polymorphism[J]. *Genetics*, 1978, **88**(1):149-169.
- [10] Li C C. The stability of an equilibrium and the average fitness of a population[J]. *Am Naturalist*, 1953, **89**(2):281-296.
- [11] Mandel S P H. The stability of a multiple allelic system[J]. *Heredity*, 1959, **13**(3):289-302.
- [12] Merrell D J. *Ecological Genetics*[M]. London: Longman, 1981. 45-98.
- [13] Ohta T, Kimura M. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population[J]. *Genet Res*, 1973, **22**(2):201-204.
- [14] Spiess E B. *Genes in Populations*[M]. New York: John Wiley & Sons, 1983. 23-145.
- [15] Wallace B. *Basic Population Genetics*[M]. New York: Columbia University Press, 1980. 49-88.
- [16] Wright S. The genetical structure of population[J]. *Ann Eugen*, 1951, **15**(3):323-354.

用超显性模型估计遗传多态平衡的种群参数

周厚高¹ 朱军¹ 闫新甫²

(¹ 广西大学农学院 南宁 530005; ² 浙江大学农学系 杭州 310029)

摘要: 基于超显性模型, 采用数学方法阐述了种群遗传多样性的延续机制及平衡条件. 从双等位基因导出的定律有较大的局限性. 在许多教科书中常用的估计平衡态基因频数的定律只适用于双等位基因. 本文用另一种方法导出一些公式, 将其扩展到多等位基因. 讨论了种群中基因数 n , 遗传负荷 (L), 杂合频 YX (H_e) 和纯合频率 (H_{om}) 之间的关系.

关键词: 超显性模型; 遗传多态平衡; 基因频 YX ; 平均适合度; 遗传负荷